图书情报工作
第64卷 第6期 2020年3月

古籍数字化研究进展述评(2009－2019)*

■ 李明杰 张纤柯 陈梦石
武汉大学信息管理学院 武汉 430072

摘 要:［目的/意义］对古籍数字化研究历程和现状进行梳理,以探寻古籍数字化研究的发展趋势。［方法/过程］运用 CiteSpace 软件对近 10 年来古籍数字化研究的作者、关键词进行共现图谱分析,分析期刊论文研究及国家社科基金项目的主要研究领域和主题分布,对现有古籍数字化产品的主要功能进行考察和体验。［结果/结论］总结出我国近 10 年来古籍数字化研究的 3 大总体特征和 5 个主要问题,指出在学术传承、用户需求、信息技术发展的驱动下,面向知识服务是未来古籍数字化研究的必然趋势。

关键词:古籍数字化 知识服务 学术传承 用户需求 技术发展
分类号:G255.1
DOI:10.13266/j.issn.0252－3116.2020.06.015

1 引言

自 20 世纪 70 年代末开始,古籍数字化研究与实践迄今已走过了近 40 年的历程。从最初的文本录入、索引编制,到大规模建设各种类型的古籍数据库(如书目型、全文型和影像型数据库等),发展到今天利用计算机信息处理技术对古籍数字文本进行各种深加工,如自动标点、自动校勘、自动注释、语义标引、文本聚类、知识挖掘等,人们对古籍数字化的认识也在不断发展。最初人们对古籍数字化的理解只是将纸质古籍的文本信息转换成数字形式,以便长久保存和利用,进而利用计算机处理信息的技术优势,开发出便捷高效的全文检索功能。随着数字信息技术的发展,特别是本世纪初“数字人文”理念的兴起,人们发现还可以将转换后的古籍数字文本当作进一步加工整理的数据,通过一系列数据处理及加工技术开发出新的知识产品,古籍数字化的概念向着更深的“知识域”迈进。当古籍数字化发展到新阶段后,就古籍整理与研究者而言,对古籍完成数字文本转换之后的海量数据进行深度的知识加工,成为绕不开的课题。为此,笔者拟对近 10 年来古籍数字化研究进行总结,分析古籍数字化研究的总体特征、存在问题及内在驱动力,并结合人文科学研究对古籍数字化产品功能需求的特点,探讨古籍数字化未来的发展方向,为古籍整理和人文社会科学相关研究提供参考和借鉴。

2 古籍数字化研究概况(2009－2019)

2.1 相关的期刊论文基本情况

笔者利用 CNKI 数据库,来源选择为“期刊”,字段选为“主题”,一次检索词为“古籍”“典籍” “善本”“古文”“古代文献”“古籍整理”“古籍保护”,二次检索词为“数字化”“电子化”“计算机”“数据库”,组配方式均为逻辑或,时间限定在 2009 年至 2019 年,勾选同义词扩展,检索日期为 2019 年 7 月 1 日,共命中 1 053 条结果。通过筛选、剔除后共得到 759 条相关结果,然后运用 CiteSpace 软件工具对所得结果进行计量和可视化分析,得到作者合作网络图和关键词共现图,见图 1 和图 2。

作者合作网络图能够反映领域内重要的研究贡献者及研究者之间的联系。在作者合作网络图中,节点的大小代表作者的发文数量。从图 1 可以看出,古籍数字化研究的作者分布较为分散,说明近 10 年来该主题研究者的来源背景比较复杂,研究的方向也较为分散;从发文数量来看,葛怀东、毛建军、张文亮等学者位居前列;还有部分研究者是以合作的方式进行学术探

* 本文系国家社会科学基金项目“文献整理学术传统在古籍数字化中的价值实现研究”(项目编号:17BTQ009)研究成果之一。
作者简介:李明杰(ORCID:0000－0002－1876－9040),教授,博士生导师,E-mail:limingjie@whu.edu.com;张纤柯(ORCID:0000－0001－6309－5682),硕士研究生;陈梦石(ORCID:0000－0002－6479－2137),硕士研究生。
收稿日期:2019－07－04 修回日期:2019－11－28 本文起止页码:130－137 本文责任编辑:徐健

李明杰，张纤柯，陈梦石. 古籍数字化研究进展述评(2009－2019)[J]. 图书情报工作,2020,64(6):130－137.

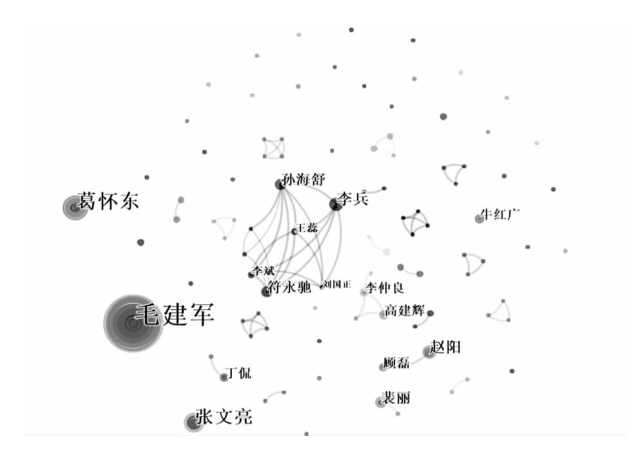


图 1　古籍数字化研究的作者合作网络

索，例如李兵、符永驰、王蕊等，这些学者均来自同一研究机构，且研究内容多以中医药古籍为主。分散的作者网络图从侧面说明了两个问题：①研究者来自于不同的学科背景，印证了现阶段古籍数字化研究的多学科性质，涉及到图书情报、文史、计算机领域，但不同类型的研究机构之间的合作较少，学科交叉的性质仍有待加强；②研究者正在努力发掘新的课题，因而研究主题显得较为分散，当前并未形成一个稳定、完整的研究体系。

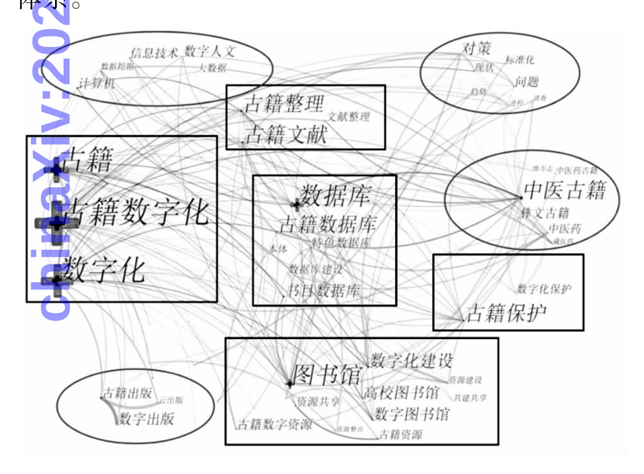


图 2　古籍数字化研究的关键词共现

关键词体现出文章所要表达的核心主题和主要内容。通过关键词共现分析，可以对研究领域的主题进行分析。在关键词共现图中，节点的大小代表关键词的频次，关键词出现频次越多，对应的节点就越大。节点之间的连线则表示共现关系，其粗细表示的是共现强度。图 2 将共现次数较多的关键词和一些特殊的关键词分别用方形框和椭圆框标记了出来。结合图 2 及调研相关文献，笔者归纳出近 10 年来古籍数字化研究主要集中在以下 6 个方面：

（1）古籍数字化概念及古籍数字化现状研究。此类论文数量偏多，说明学界比较关注古籍数字化的基本理论问题，同时注重与实践相结合。对于古籍数字化概念的探讨始于 20 世纪 90 年代中后期，至今没有间断。人们对古籍数字化的认识，随着古籍数字化的发展而不断深入，这个过程大致经历了古籍存储介质的转换、文本组织与检索、数据加工与知识服务 3 个阶段。对古籍数字化现状的研究，多是对古籍数字化进程、问题、对策、趋势等方面的总结和分析，代表性论文有《中文古籍数字化的现状与意义》[1]《我国古籍数字化进展与研究述评》[2]《中国大陆地区古籍数字化问题及对策》[3]《古籍数字化的困局及应对策略》[4]等。

（2）与古籍保护相关的古籍数字化研究。视古籍数字化为古籍再生性保护的一种有效手段，并展开相应的研究，代表性的论文有《古籍保护与开发的策略与建议》[5]《古籍数字化的保真问题》[6]《数字化生存下的历史文化资源保护与开发研究——以陕西为中心》[7]《数字人文时代图书馆古籍文献资源的数字化长期保存机制探讨》[8]等。

（3）与图书馆业务相关的古籍数字资源建设研究。主要涉及各类型古籍数据库的开发、古籍数字资源整合、古籍数字资源共建共享等问题。图书馆作为古籍收藏、管理、开发与利用的主体，与其古籍数字化工作相关的研究主要围绕古籍数字资源建设展开，代表性论文如《古籍书目数据库的标准与评价研究》[9]《我国古籍书目数据库建设标准规范探讨》[10]《古籍数字资源库发展趋势探讨》[11]《古籍数字化与共建共享》[12]《图书馆古籍数字资源整合研究》[13]《古籍资源聚合研究》[14]等。另外，古籍版本数据库和专题知识库的建设也引起了学者们的关注，如《基于 Unicode 的中国蒙古文古籍版本数据库的构建》[15]《论古籍版本数据库建设》[16]《中医古籍知识库的发展现状及古籍养生知识库的构建》[17]等。

（4）专类古籍的数字化研究。涉及中医药古籍、少数民族古籍、地方文献、谱牒、方志、拓片等的数字化问题，如《中医古籍数字化整理方案研究》[18]《数字化家谱的研究与实现》[19]《抄写本方志古籍数字化整理与实践》[20]《国家图书馆石刻拓片的数字化》[21]《蒙古文资源数字化共建共享保障体系研究》[22]《中医古籍数字化建设及其平台设计和实现》[23]等。这类研究多是从文献保护和文化传承的视角展开的，就古籍数字

化技术本身而言，并无多少突破。

（5）古籍数字化协作机制研究。选题重复、缺少规划等问题是长期困扰古籍数字化的一个问题，已有不少学者意识到，必须通过国家层面的机制创新来解决这一难题。这方面代表性的论文有《对现阶段古籍数字化的思考——兼论古籍数字化国家行为的重要性》[24]《古籍数字化工作统筹协调机制的构建》[25]《试论古籍数字化的国家宏观政策体系》[26]《中文古籍数字化的主体构成及协作机制初探》[27]等。

（6）古籍数字化与数字人文相结合的研究，包括关联数据、知识挖掘、地理信息系统（GIS）、知识组织等。近5年来，这类论文有逐年增多的趋势，预示着古籍数字化研究已进入面向知识服务的新阶段。代表性的论文如《命名实体识别在方志内容挖掘中的应用研究》[28]《"数字人文"内涵与古籍数字化的深度开发》[29]《"数字人文"与历史地理信息化研究》[30]《基于GIS的中国古籍地理信息系统研究》[31]《面向数字人文研究的大规模古籍文本可视化分析与挖掘》[32]《地方历史文献的数字化、数据化与文本挖掘：以中国地方历史文献数据库为例》[33]《基于本体和XML的注疏文献的结构化知识表示》[34]《中国历史地理关联数据在图书馆数字人文项目中的开放应用研究》[35]等。在中医古籍的数字化研究方面，对知识的挖掘和重组取得了令人瞩目的成果，如《基于本体论构建中医古籍知识库的探索》[36]《基于知识元的中医古籍方剂知识表示研究》[37]《基于温病古籍知识概念语义网络构建研究》[38]《基于本体的中医学术传承脉络构建》[39]等。

2.2 相关的国家社科基金项目基本情况

笔者利用国家社会科学基金项目数据库[40]，将时间范围设定在2009－2018年，学科类别限定在"图书馆、情报与文献学"，检得与古籍数字化相关的研究项目54项。从立项数量来看，整体呈逐年上升趋势，其中2009年最少，只有1项；2017年最多，达到13项。从项目类型来看，西部项目最多，为23项，占比42.59%；其次是一般项目21项，占比38.89%；青年项目9项，占比16.67%；重大项目1项，占比1.85%。从项目研究的主题分布来看，大致分为以下几类：

（1）少数民族古籍和中医药古籍的抢救性整理与保护。此类项目数量最多，计有21项，典型的如"布依族象形文字古越书价值及其资源抢救、建设与开发利用研究"（10XTQ010）、"蒙古民族文献价值及其文献资源建设研究"（10XTQ012）、"云南傣族医药古籍文献整理及其基础数据库建设研究"（11CTQ041）、"撒拉族文献资料数据库的建设研究"（11XTQ011）、"四川南派藏医药古籍文献的抢救性挖掘整理及数据库研究"（13XTQ012）、"青海地区藏医药古籍文献收集、整理与数字化研究"（16BTQ051）、"湘黔桂边区侗医药古籍文献收集、整理及数字化研究"（17CTQ017）、"存世宁夏旧方志数字化整理研究"（18XTQ007）等。

（2）口述历史、特殊档案及非物质文化遗产保护相关的数字化项目。这类项目严格地说不属于古籍数字化项目，但属历史文献的范畴，故放在一起统计。此类项目计有15项，如"云南少数民族口述文献价值及其数字化语音库建设"（10CTQ015）、"白族口承文艺非物质文化遗产调查及专题数据库建设"（12CTQ018）、"历代孔府档案文献集成与研究及全文数据库建设"（13&ZD108）、"清代彝事汉文硃批档案文献汇编、整理与数字化建设"（16BTQ045）、"濒危口头诗学青海撒拉族'玉尔'的数字化保护与传承研究"（17XTQ011）、"濒危非物质文化遗产的青海土族'道拉'的传承与数字化保护研究"（18BTQ035）等。自2016年起，国家社会科学基金项目明显加大了对少数民族地区口述历史和非物质文化遗产的保护力度。

（3）专题研究或专类古籍的数字化项目。此类项目计有14项。在专题研究方面，有的探讨古籍数字化理论问题，如"文献整理学术传统在古籍数字化中的价值实现研究"（17BTQ009）、"图书馆古籍文献的数字人文开发与应用模式研究"（17XTQ003）等；有的探讨古籍数字化产业发展问题，如"我国台湾地区数位典藏产业链建置研究17CTQ033"；有的致力于某类古籍的数字化平台建设，如"越南与中国南海地名考证及历史地理信息平台建设研究"（18BTQ048）；有的探讨古籍数字化的具体技术问题，如"面向数字出版的蒙古文字形规范问题研究"（17XTQ001）、"基于典籍的中华传统文化知识表达体系自动构建方法研究"（18BTQ063）、"徽州文书分类法与元数据标准设计研究"（12CTQ007）、"基于关联数据的南海水下文化遗产文献资源共享和可视化检索研究"（17CTQ004）、"基于方志类典籍的地标农产品挖掘与组织研究"（18CTQ011）等。专类古籍的数字化通常是为某一领域的专题研究服务的，具有很强的应用性，如"四库全书研究资源数据库建设"（09BTQ017）、"青海省三江源区藏传佛教寺院特色数

据库建设研究”(10XTQ013)、“湖南碑刻整理研究和数据库建设”(11XTQ010)、“民国以前云南水资料整理及数据库建设”(13BTQ041)、“存世宁夏旧方志数字化整理研究”(18XTQ007)。

(4)单书的古籍数字化研究。此类项目计有 4 项，如“楚辞文献语义化研究”(10BTQ031)、“基于《苯教大藏经・丹珠尔》的词汇计量研究”(12CTQ020)、“《秦简书法大字典》创编及数字化应用研究”(12XTQ009)、“《读史方舆纪要》地名考证及地理信息系统平台建设研究”(17BTQ038)。

3 古籍数字化研究的总体特征和存在问题

3.1 总体特征

通过对近 10 年来古籍数字化领域期刊论文发表和国家社科基金项目立项情况的分析，笔者认为，近 10 年来古籍数字化研究总体上呈现以下三大特征：

(1)古籍的再生性保护仍是当前古籍数字化实践和研究的主要领域。从论文发表和项目立项情况来看，专类古籍保护和图书馆古籍资源建设类的古籍数字化研究占了绝大多数。换句话说，当前我国古籍数字化仍是以保护和传承古籍的内容文本为主要任务，遵循的是古籍的再生性保护思路。这在那些具有浓郁的民族特色、地区特色、文化特色的古籍身上体现得尤为突出。国家社科基金项目中与古籍数字化相关的“西部项目”数量最多，除了政策倾斜的因素之外，更多的是因为西部地区这 3 类特色古籍较多，亟需得到保护。仅从技术上而言，此类研究也许谈不上有多大创新，但从文献保护和文化传承的角度而言，则意义非同一般。

(2)古籍数字化研究领域不断拓展，学科化趋势日益明显。如果把古籍数字化的研究内容整体比作一座金字塔，那么古籍再生性保护就是这座金字塔的塔基。在塔基之上，不断延展出新的研究领域，比如古籍数字化基础理论、古籍数字资源建设、古籍数字资源管理、古籍数字化政策、古籍数字化技术等。从理论到实践，从技术到政策，从应用到管理，古籍数字化研究俨然可以构成一门新的学科。事实上，古籍数字化的学科建设问题已经引起了部分学者的关注，如毛建军从传统目录学、版本学、校勘学研究的新视野，探讨了电子文献学的研究范畴[41]；葛怀东对古籍数字化学科的研究对象、学科性质作了初步界定，并提出拟以数字化古籍整理学、中文信息处理技术、古籍数字化管理理论作为学科构建的基本框架[42]。

(3)古籍数字化研究在技术和理念上形成了以知识服务为核心的思想。古籍数字化不论如何发展，其本质仍属于古籍整理的范畴。传统的古籍整理是为人文社会科学研究提供文本和语义服务，在数字化时代，古籍整理进入了更深的知识层次。在古籍数字化这座金字塔的塔尖，古籍数字化技术与理念指引着古籍数字化发展的方向。近 10 年来，数字技术与人文观念的结合，催生了以知识服务于人文社会科学研究的思想，使得古籍数字化整理不再局限于提供准确可靠、可理解的内容文本，而是提供可用于辅助研究的知识产品。

3.2 存在问题

除了研究论文与科研项目之外，具体的数字化产品也是古籍数字化研究成果的一部分。因此，笔者在对研究论文和科研项目进行统计分析之外，还通过网络调查、亲身体验等方法，考察了现有古籍数字化成果的系统功能。将文献调研与系统功能考察结合起来，笔者总结出古籍数字化研究存在以下不足：

(1)忽视了对用户需求的调查分析。古籍数字化成果最终面向的是人文社会科学研究者及古籍爱好者群体。古籍数字化除了运用技术手段转录古籍内容、保护古籍本身外，更重要的是服务学术研究。当前的古籍数字化成果究竟能在多大程度上满足传统学术研究的需要，是很值得怀疑的，因为信息技术专家对人文学科研究的特点及专业知识信息需求缺乏深入的调查分析。

(2)未能构建完整的古籍数字化学术规范。古籍数字化不是古籍内容存储介质的简单转换，而是传统的古籍整理在数字化环境下的合理延伸。古籍整理的理念、方法经过长时期的历史沉淀，适应了传统学术研究的需要，有其科学合理的一面，如对底本的考订和遴选、对文字的校勘、对副文本的处理等。因此，古籍数字化产品必须是古籍整理的成果，必须符合古籍整理的学术规范。然而时至今日，古籍数字化在文本加工、技术处理、管理流程等方面，还没有建立一套相应的规范来保障古籍数字化的学术品质。

(3)对古籍整理学术传统的借鉴力度不够，学科合作有待加强。我国的古籍整理经过漫长的历史积累，形成了许多优良的学术思想，如孔子提出的“述而

不作”的文献整理基本原则、司马迁提出的文献编纂的“会通观”、支敏度提出的“百虑一致，万流同归”的版本源流考订思想、颜之推提出的“观天下书未遍，不得妄下雌黄”的校勘学思想、章学诚提出的“辨章学术，考镜源流”的目录学思想等。这些学术思想是与古籍整理相适应的，对今天的古籍数字化实践仍具有重要的借鉴意义。

（4）忽视了对现有古籍整理成果的数字化研究。在纸本文献环境下，前人已经做了大量的古籍整理工作，留下了丰富的古籍整理成果，包括大量的古籍序跋、书目、类序、提要以及编纂而成的各种工具书，如职官表、年表、传记、年谱、学案、类书等。这些成果本来就是用以辅助传统学术研究的，从提高服务效率的角度来讲，对已有古籍整理成果的数字化将起到事半功倍的效果。然而，现有的古籍数字化产品较少将这类古籍整理成果嵌入其中，造成了学术资源的浪费。

（5）案例研究所涉古籍范围有限。当前的古籍数字化研究已经进入数据整理阶段，本体的构建、语义的关联、知识的挖掘和可视化技术，已大量应用到古籍数字化领域。但从应用的古籍范围来看，主要局限于医药、农业、家谱类古籍，而像历代笔记小说、各家书目、类书、方志等其他有特点的古籍类别还较为少见，这需要在今后的研究中加以拓展。另外，从实践来看，对古籍知识深度开发的成果还不是很多，尤其缺乏集各种功能于一体的开放共享式的古籍数字化平台。

4 面向用户的知识服务：未来古籍数字化研究趋势探讨

如前所述，古籍再生性保护是当前古籍数字化这座金字塔的塔基，折射的是古籍数字化的现实情况，而古籍数字化技术与理念则是这座金字塔的塔尖，指引的是古籍数字化发展的未来方向。当现有的各大图书馆收藏的珍本、孤本及普通古籍，以及地方特色的少数民族古籍都完成了数字文本的转换（即再生性保护）之后，对古籍数字资源的深加工就成为未来古籍数字化研究的必然课题。就目前古籍数据加工的技术和理念来看，面向用户的知识服务是未来古籍数字化研究的趋势，而这也是由学术传承、用户需求及信息技术发展为内在驱动力的。

4.1 学术传承驱动下的知识服务

中华文化最大的一个特性在于其传承性。对于古籍而言，一方面它承载的知识内容是历代学术研究的成果，具有传承性；另一方面，古籍本身是历代学者整理和出版的产物，在长期实践中形成的文献整理思想和方法体系，也是学术传承的一部分。对于古籍数字化而言，一方面要保存古籍的知识内容，传承历代学术研究的成果，另一方面要继承和发扬古代文献整理的优良学术传统。

而从知识生产的逻辑来看，任何学术研究都肩负着知识生产、学术创新的任务，而在推进知识生产的过程中，研究者们最先做的工作都是文献回顾。它指的是对到目前为止的、与某一研究问题相关的各种文献进行系统查阅和分析，以了解该领域研究状况的过程。或者说，就是一个系统识别、寻找、考察和总结那些与我们的研究有关的文献的过程[43]。文献回顾需要在既有文献和新的知识生产之间建立内在的学术联系：一是通过清理既有文献，为新的知识生产过程建立一个学术对话点；二是通过文献清理，发现新的研究空间，从而建立一个学术推进点[44]。它除了推动知识生产和学术创新外，另一个关键点在于理清学术源流、促进学术传承。在传统文献环境下，古人通过对前代文献的系统整理来完成对前代学术的总结，几成惯例，如西汉刘向等学者对政府藏书的整理，通过“条别篇目，撮其指意”，撰成《别录》；刘歆类分群书，编成《七略》，从而完成对先秦以来学术的系统总结。历代大型类书（如《永乐大典》）和综合性丛书（如《四库全书》）的编纂，也是对一代学术的系统梳理和集成。

在数字环境下，现代学者对承载古人思想和智慧的古籍文本进行存储介质的转换、语义的解析、内容的组织、知识的挖掘等工作，与古人对前代文献的系统整理的本质是一脉相承的，只是过程和方法上有所不同而已。在对古籍文本内容进行整理的同时，古籍整理的学术传统也在古籍数字化中得以传承，如在古籍文本的转录过程中，对文字的原始特征及各种眉批、夹住、点评予以保留，体现了“述而不作”的原则；对不同类型的古籍数字资源及各种辅助人文学科研究功能的集成，体现了“会通”的理念；而通过语义分析和关联，实现文本的自动聚类，甚至利用知识挖掘的方法，揭示古籍文本中隐含的丰富知识和不同的学科主题内容，理清各家思想，探究学术源流，秉承了“辨章学术，考镜源流”的思想。

无论是古代的文献生产与整理活动，还是当代信

李明杰，张纤柯，陈梦石. 古籍数字化研究进展述评(2009－2019)[J]. 图书情报工作,2020,64(6):130－137.

息环境下的古籍数字化实践，从本质上讲都是知识生产与传播的中间环节。知识生产的主体根据自身的社会实践经历，将个人经验予以提炼和升华，输出知识成果。但经过一段时间的传播后，由于传播过程中的不可避免的信息失真，后人必须对前人的知识成果进行整理，发掘和吸纳其可利用的成分，从而产生新的认知和研究主题。如此循环往复，构成了以知识为核心的学术传承过程。

4.2 用户需求驱动下的知识服务

回顾古籍数字化研究的发端到全面开展的过程，再结合近10年来的研究现状，笔者发现，服务于工作实践、解决实际问题、满足用户在学术研究中的知识信息需求一直是古籍数字化的核心议题。从古籍收藏的主要机构——图书馆来看，其服务方式经历了从文献、信息到知识的转变，而古籍数字化的发展也同样经历了古籍文本的数字化保存、文本信息的组织加工、文本内知识的挖掘和再组织的过程，这个不断递进的过程也是用户需求不断发展的结果。

为了验证用户需求对面向知识服务的古籍数字化的驱动作用，徐丹[45]将古籍数字化产品功能分为"基本信息浏览""内容检索与筛选""阅读显示""辅助工具""个性化设置"和"知识增值利用"6个功能模块及具体的33个功能点，利用线下访谈、线上问卷的方式，对人文社会科学研究者对古籍数字化产品功能的需求，以及23所大型公共图书馆和高校图书馆使用的古籍数字化产品功能进行了调查，然后引入日本狩野纪昭教授的KANO模型，对各个模块的功能与用户满意度之间的关系进行评测。结果表明，"基本信息浏览""内容检索与筛选"和"阅读显示"这3个功能模块属于基本需求和期望型需求，而"知识增值利用"属于魅力型需求。"知识增值利用"功能包括联机字典、纪年换算、词/字频统计、社会网络分析、文本聚类、知识关联、内容可视化、地理空间展示等具体功能。古籍数字化产品如果能提供这些功能，将会大大提升用户的满意度。目前，古籍数字化产品在"基本信息浏览""内容检索与筛选"和"阅读显示"功能方面已经基本实现，随着技术的发展、研究的深入，"知识增值利用"这一部分功能逐渐从用户的魅力型需求转向期望型需求或者基本型需求。因此，对古籍文本转换后的数据进行深度知识挖掘、加工和可视化展示，为用户提供知识服务，将是古籍数字化研究的重点内容。

4.3 信息技术发展驱动下的知识服务

信息技术作为一种工具，在科学认知、科学研究中发挥着重要的功能。它不仅能带来研究效率的提高，还能拓宽研究思路，尤其是数字技术与传统学术研究相结合，甚至带来研究方法的变革，数字人文就是最好的例子。古籍数字化既是关于古籍整理的研究，也是关于信息技术应用的研究。只有把握好信息技术的发展导向，才能认清古籍数字化研究的未来趋势。

从信息技术的发展历程来看，古籍数字化经历了古籍文本的数字化转换、古籍内容的组织加工、古籍知识的发掘和可视化呈现3个阶段。第一阶段所用的技术方法包括文字录入、影像扫描、OCR识别等，其中也涉及汉字的简繁转换、俗体字和异体字的识别、字符集标准等技术问题；第二阶段所用的技术方法包括运用计算机进行分类标引、书目编制、关键词索引，以及自动标点、校勘、注释、翻译、编纂等，主要满足用户对特定主题或学科领域内的古籍原文内容及语义阐释的需求。现阶段，随着信息技术的进一步发展，古籍数字化越来越重视古籍知识的发掘和可视化呈现，这种以海量的古籍文本数据为对象的知识加工服务，打破了学科与学科、书与书之间的界限，技术上主要采用文本聚类、数据挖掘、本体构建、社会网络分析、地理信息系统、知识图谱等方法。

从目前情况来看，面向知识服务的信息技术主要应用于中医药、农业类的古籍数字化，如《胸痹心痛古代文献用药规律及聚类分析》[46]《基于古籍文献挖掘的头痛中医病性证素特征研究》[47]《基于中医古籍数据挖掘的不育症病名源流考辨》[48]《基于数据挖掘技术分析古籍中风病腧穴处方规律》[49]《基于本体的脾脏象理论知识体系构建研究》[50]《基于中医骨伤古籍本体的语义检索研究》[51]《农业古籍本体构建及应用》[52]《农业古籍本体的构建及其检索机制研究》[53]等。另外，在方志、地理、史书类古籍方面，也有不少成果，如《方志类古籍中物产名与别名关系的可视化——基于社会网络分析技术视角》[54]《基于GIS的中国古籍地理信息系统研究》[55]《大规模古籍文本在中国史定量研究中的应用探索》[56]等。其他领域面向知识服务的古籍数字化研究还有很大的空间。

目前，古籍数字化在古籍知识价值的拓展阶段，出现了由古籍数字资源库向知识共享平台转变的新动向。这些知识共享平台，除了整合大量的古籍数字资

源、提供浏览和检索功能之外,还建立了相应的知识社区,提供知识分享、主题讨论等功能。如“籍合网”创建了线上知识社区“古籍圈”,用户在圈内可以自由提问、发言、组成团体,对相关问题进行讨论,互助解决问题,还可以分享学术随便、研究心得。“中国哲学书电子化计划(CTP)”除了提供原典资料库、原典影印本、内部字典、高级搜索功能外,还提供历代注释本显示、当代研究资料库、后设资料系统、语词分析表、相似段落资料等高级功能,同时可以在讨论区提问和参与相关问题讨论,在维基区共同编辑、纠正古籍全文中的错字[52]。

5 结语

古籍数字化学术研究是古籍数字化实践活动的先导。通过对近 10 年来发表的与古籍数字化相关的学术论文及国家社科基金项目的研究领域、主题分布的统计分析,结合对现有古籍数字化产品功能的考察和亲身体验,笔者归纳总结了近 10 年来古籍数字化研究的总体特征和存在的问题,并从学术传统的传承、用户需求的满足、技术发展的推动 3 个方面,指明面向用户的知识服务是未来古籍数字化研究的趋势。古籍数字化在做好古籍文化遗产保护、数字资源建设工作的同时,应根据人文社会科学研究的特点,面向专业研究者的知识需求,开发出更多辅助人文社会科学研究的知识产品。

参考文献:

[1] 刘伟红. 中文古籍数字化的现状与意义[J]. 图书与情报,2009(4):134-137.
[2] 朱锁玲,包平. 我国古籍数字化进展与研究述评[J]. 图书馆理论与实践,2009(9):18-21.
[3] 高娟,刘家真. 中国大陆地区古籍数字化问题及对策[J]. 中国图书馆学报,2013(4):110-119.
[4] 邵正坤. 古籍数字化的困局及应对策略[J]. 图书馆学研究,2014(12):32-34.
[5] 刘家真,程万高. 古籍保护与开发的策略与建议[J]. 中国图书馆学报,2009(3):15-20.
[6] 丁侃,柳长华. 古籍数字化的保真问题[J]. 中医文献杂志,2009(2):31-33.
[7] 赵东. 数字化生存下的历史文化资源保护与开发研究——以陕西为中心[D]. 济南:山东大学,2014.
[8] 陈诚. 数字人文时代图书馆古籍文献资源的数字化长期保存机制探讨[J]. 中国中医药图书情报杂志,2019(3):6-9.
[9] 毛建军. 古籍书目数据库的标准与评价研究[J]. 图书馆理论与实践,2009(6):30-33.
[10] 周琳洁. 我国古籍书目数据库建设标准规范探讨[J]. 图书馆建设,2010(2):47-50.
[11] 赵葆英,苏沫. 古籍数字资源库发展趋势探讨[J]. 图书馆工作与研究,2011(2):36-39.
[12] 梁爱民,陈荔京. 古籍数字化与共建共享[J]. 国家图书馆学刊,2012(5):108-112.
[13] 萨雷. 图书馆古籍数字资源整合研究[J]. 图书馆,2014(4):116-119.
[14] 陈士兵. 古籍资源聚合研究[D]. 太原:山西大学,2018.
[15] 黄飞龙,札义兰. 基于 Unicode 的中国蒙古文古籍版本数据库的构建[J]. 内蒙古民族大学学报,2009(3):160-161.
[16] 葛怀东. 论古籍版本数据库建设[J]. 情报探索,2012(7):75-77.
[17] 安欢等. 中医古籍知识库的发展现状及古籍养生知识库的构建[J]. 西部中医药,2016(6):158-159.
[18] 李兵. 中医古籍数字化整理方案研究[D]. 北京:中国中医科学院,2009.
[19] 柯洲. 数字化家谱的研究与实现[D]. 武汉:华中师范大学,2011.
[20] 胡以涛,宋叶. 抄写本方志古籍数字化整理与实践[J]. 图书馆理论与实践,2014(8):101-103.
[21] 袁玉红. 国家图书馆石刻拓片的数字化[J]. 图书馆理论与实践,2014(5):17-18.
[22] 王康,康健. 蒙古文资源数字化共建共享保障体系研究[J]. 图书情报工作,2015,59(12):39-43.
[23] 曹霞,常存库,裴丽. 中医古籍数字化建设及其平台设计和实现[J]. 中华医学图书情报杂志,2016(3):45-47.
[24] 李玉安. 对现阶段古籍数字化的思考[C]//第二届中国古籍数字化国际学术研讨会论文集. 北京:五洲传播出版社,2011.
[25] 陈得媛. 古籍数字化工作统筹协调机制的构建[J]. 中国图书馆学报,2009(5):47-51.
[26] 郜丽红. 试论古籍数字化的国家宏观政策体系[J]. 中州大学学报,2010(4):39-41.
[27] 李明杰,俞优优. 中文古籍数字化的主体构成及协作机制初探[J]. 图书与情报,2010(1):34-44.
[28] 朱锁玲. 命名实体识别在方志内容挖掘中的应用研究[D]. 南京:南京农业大学,2011.
[29] 范佳. “数字人文”内涵与古籍数字化的深度开发[J]. 图书馆学研究,2013(3):29-32.
[30] 陈刚. “数字人文”与历史地理信息化研究[J]. 南京社会科学,2014(3):136-142.
[31] 王大学,陈熙,杨光辉. 基于 GIS 的中国古籍地理信息系统研究[J]. 复旦学报(自科版),2016(6):684-688.
[32] 欧阳剑. 面向数字人文研究的大规模古籍文本可视化分析与挖掘[J]. 中国图书馆学报,2016(2):66-80.
[33] 赵思渊. 地方历史文献的数字化、数据化与文本挖掘:以中国

地方历史文献数据库为例[J]. 清史研究,2016(4):26－35.
[34] 马创新,陈小荷. 基于本体和XML的注疏文献的结构化知识表示[J]. 图书馆杂志,2017(8):62－68.
[35] 夏翠娟. 中国历史地理关联数据在图书馆数字人文项目中的开放应用研究[J]. 中国图书馆学报,2017(2):40－53.
[36] 孙海舒,符永驰,张华敏,等. 基于本体论构建中医古籍知识库的探索[J]. 医学信息学杂志,2011(3):64－68.
[37] 丁侃. 基于知识元的中医古籍方剂知识表示研究[D]. 北京:中国中医科学院,2012.
[38] 许雯. 基于温病古籍知识概念语义网络构建研究[D]. 北京:中国中医科学院,2015.
[39] 丁侃,张丽君. 基于本体的中医学术传承脉络构建[J]. 中医文献杂志,2018(5):32－35.
[40] 全国哲学社会科学工作办公室. 国家社科基金项目数据库[DB/OL]. [2019－04－19]. http://fz.people.com.cn/skygb/sk/.
[41] 毛建军. 论古籍电子文献学研究范畴的确立[J]. 图书馆理论与实践,2010(9):46－48,88.
[42] 葛怀东. 古籍数字化的学科建设[J]. 科技教育创新,2012(1):156－157.
[43] 风笑天. 论社会研究中的文献回顾[J]. 华中师范大学学报(人文社会科学版),2010(4):40－46.
[44] 张兆曙,高远欣. 知识生产与文献回顾——从技术指引到意义指引[J]. 天津社会科学,2019(1):48－53.
[45] 徐丹. 基于KANO模型的人文学科研究需求对古籍数字化产品的功能调查[D]. 武汉:武汉大学,2019.
[46] 陆一竹,李园白,毛静远,等. 胸痹心痛古代文献用药规律及聚类分析[J]. 天津中医药,2013(10):629－631.
[47] 雷玉娇. 基于古籍文献挖掘的头痛中医病性证素特征研究[D]. 郑州:河南中医药大学,2018.
[48] 邢益涛,张明强,王定国,等. 基于中医古籍数据挖掘的不育症病名源流考辨[J]. 河南中医,2018(4):603－605.
[49] 陈楚云,李丽霞,李伟林,等. 基于数据挖掘技术分析古籍中风病腧穴处方规律[J]. 上海针灸杂志,2017(12):1494－1499.
[50 崔家鹏,王彩霞,袁东超,等. 基于本体的脾脏象理论知识体系构建研究[J]. 中华中医药学刊,2018(2):388－292.
[51] 林尔正. 基于中医骨伤古籍本体的语义检索研究[D]. 福州:福建中医学院,2018.
[52] 曹玲,何琳. 农业古籍本体构建及应用[J]. 广西师范大学学报(自科版),2007(2):1－4.
[53] 何琳,曹玲. 农业古籍本体的构建及其检索机制研究[J]. 现代图书情报技术,2006(12):37－53.
[54] 李娜,包平. 方志类古籍中物产名与别名关系的可视化——基于社会网络分析技术视角[J]. 图书馆论坛,2017(12):108－114.
[55] 王大学. 基于GIS的中国古籍地理信息系统研究[J]. 复旦学报(自科版),2016(6):684－697.
[56] 欧阳剑. 大规模古籍文本在中国史定量研究中的应用探索[J]. 大学图书馆学报,2016(3):5－15.

作者贡献说明：
李明杰：提出论文选题，设计论文结构，对重要观点进行归纳总结；
张纤柯：收集论文资料，撰写初稿；
陈梦石：完成共现图谱分析，参与初稿撰写。

Review on the Research Progress of the Digitization of Ancient Chinese Books (2009－2019)

Li Mingjie　Zhang Xianke　Chen Mengshi
School of Information Management, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] In order to explore the development trend of the research on the digitization of ancient books, it makes a systematic investigation on the research status of the digitization of ancient books in China. [**Method/process**] By using CiteSpace software, authors and key words of the research on the digitization of ancient books in the past decade were analyzed. Also, it analyzed the main research fields and subject distribution of papers and national social science fund projects. Authors made an investigation and had an experience on the functions of the existing digital products of ancient books. [**Result/conclusion**] This paper summarizes 3 general characteristics and five main problems of the research of digitization of ancient books in recent ten years, and points out that the knowledge service digitization of ancient books is the inevitable trend in the future under three drives of academic inheritance, user demand and the development of information technology.

Keywords: digitization of ancient books　knowledge service　academic inheritance　user demand　technology development